

10/552341

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/522,341A
Source: PJT
Date Processed by STIC: 4/14/06

ENTERED



PCT

RAW SEQUENCE LISTING

DATE: 04/14/2006

PATENT APPLICATION: US/10/522,341A

TIME: 10:12:52

Input Set : F:\Final sequence list-12810-00057-US.txt

Output Set : N:\CRF4\04142006\J522341A.raw

3 <110> APPLICANT: Kock, Michael
 4 Frank, Markus
 5 Badur, Ralf
 7 <120> TITLE OF INVENTION: Novel selection method
 9 <130> FILE REFERENCE: 12810-00057-US
 11 <140> CURRENT APPLICATION NUMBER: US 10/522,341A
 12 <141> CURRENT FILING DATE: 2005-01-25
 14 <150> PRIOR APPLICATION NUMBER: PCT/EP2003/007877
 15 <151> PRIOR FILING DATE: 2003-07-18
 17 <150> PRIOR APPLICATION NUMBER: DE 102 34 287.3
 18 <151> PRIOR FILING DATE: 2002-07-26
 20 <160> NUMBER OF SEQ ID NOS: 179
 22 <170> SOFTWARE: PatentIn version 3.3
 25 <210> SEQ ID NO: 1
 26 <211> LENGTH: 1284
 27 <212> TYPE: DNA
 28 <213> ORGANISM: Escherichia coli
 30 <220> FEATURE:
 31 <221> NAME/KEY: CDS
 32 <222> LOCATION: (1)..(1281)
 33 <223> OTHER INFORMATION: coding for cytosine deaminase (codA)
 35 <400> SEQUENCE: 1
 36 gtg tcg aat aac gct tta caa aca att att aac gcc cgg tta cca ggc 48
 37 Val Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly
 38 1 5 10 15
 39 gaa gag ggg ctg tgg cag att cat ctg cag gac gga aaa atc agc gcc 96
 40 Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala
 41 20 25 30
 42 att gat gcg caa tcc ggc gtg atg ccc ata act gaa aac agc ctg gat 144
 43 Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp
 44 35 40 45
 45 gcc gaa caa ggt tta gtt ata ccg ccg ttt gtg gag cca cat att cac 192
 46 Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His
 47 50 55 60
 48 ctg gac acc acg caa acc gcc gga caa ccg aac tgg aat cag tcc ggc 240
 49 Leu Asp Thr Thr Gln Thr Ala Gly Gln Pro Asn Trp Asn Gln Ser Gly
 50 65 70 75 80
 51 acg ctg ttt gaa ggc att gaa cgc tgg gcc gag cgc aaa gcg tta tta 288
 52 Thr Leu Phe Glu Gly Ile Glu Arg Trp Ala Glu Arg Lys Ala Leu Leu
 53 85 90 95
 54 acc cat gac gat gtg aaa caa cgc gca tgg caa acg ctg aaa tgg cag 336
 55 Thr His Asp Asp Val Lys Gln Arg Ala Trp Gln Thr Leu Lys Trp Gln
 56 100 105 110

P.6

RAW SEQUENCE LISTING

DATE: 04/14/2006

PATENT APPLICATION: US/10/522,341A

TIME: 10:12:52

Input Set : F:\Final sequence list-12810-00057-US.txt

Output Set: N:\CRF4\04142006\J522341A.raw

```

57 att gcc aac ggc att cag cat gtg cgt acc cat gtc gat gtt tcg gat 384
58 Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp
59      115      120      125
60 gca acg cta act gcg ctg aaa gca atg ctg gaa gtg aag cag gaa gtc 432
61 Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val
62      130      135      140
63 gcg ccg tgg att gat ctg caa atc gtc gcc ttc cct cag gaa ggg att 480
64 Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile
65 145      150      155      160
66 ttg tcg tat ccc aac ggt gaa gcg ttg ctg gaa gag gcg tta cgc tta 528
67 Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu
68      165      170      175
69 ggg gca gat gta gtg ggg gcg att ccg cat ttt gaa ttt acc cgt gaa 576
70 Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu
71      180      185      190
72 tac ggc gtg gag tcg ctg cat aaa acc ttc gcc ctg gcg caa aaa tac 624
73 Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr
74      195      200      205
75 gac cgt ctc atc gac gtt cac tgt gat gag atc gat gac gag cag tcg 672
76 Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser
77      210      215      220
78 cgc ttt gtc gaa acc gtt gct gcc ctg gcg cac cat gaa ggc atg ggc 720
79 Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly
80 225      230      235      240
81 gcg cga gtc acc gcc agc cac acc acg gca atg cac tcc tat aac ggg 768
82 Ala Arg Val Thr Ala Ser His Thr Thr Ala Met His Ser Tyr Asn Gly
83      245      250      255
84 gcg tat acc tca cgc ctg ttc cgc ttg ctg aaa atg tcc ggt att aac 816
85 Ala Tyr Thr Ser Arg Leu Phe Arg Leu Leu Lys Met Ser Gly Ile Asn
86      260      265      270
87 ttt gtc gcc aac ccg ctg gtc aat att cat ctg caa gga cgt ttc gat 864
88 Phe Val Ala Asn Pro Leu Val Asn Ile His Leu Gln Gly Arg Phe Asp
89      275      280      285
90 acg tat cca aaa cgt cgc ggc atc acg cgc gtt aaa gag atg ctg gag 912
91 Thr Tyr Pro Lys Arg Arg Gly Ile Thr Arg Val Lys Glu Met Leu Glu
92      290      295      300
93 tcc ggc att aac gtc tgc ttt ggt cac gat gat gtc ttc gat ccg tgg 960
94 Ser Gly Ile Asn Val Cys Phe Gly His Asp Asp Val Phe Asp Pro Trp
95 305      310      315      320
96 tat ccg ctg gga acg gcg aat atg ctg caa gtg ctg cat atg ggg ctg 1008
97 Tyr Pro Leu Gly Thr Ala Asn Met Leu Gln Val Leu His Met Gly Leu
98      325      330      335
99 cat gtt tgc cag ttg atg ggc tac ggg cag att aac gat ggc ctg aat 1056
100 His Val Cys Gln Leu Met Gly Tyr Gly Gln Ile Asn Asp Gly Leu Asn
101      340      345      350
102 tta atc acc cac cac agc gca agg acg ttg aat ttg cag gat tac ggc 1104
103 Leu Ile Thr His His Ser Ala Arg Thr Leu Asn Leu Gln Asp Tyr Gly
104      355      360      365
105 att gcc gcc gga aac agc gcc aac ctg att atc ctg ccg gct gaa aat 1152

```

RAW SEQUENCE LISTING

DATE: 04/14/2006

PATENT APPLICATION: US/10/522,341A

TIME: 10:12:52

Input Set : F:\Final sequence list-12810-00057-US.txt

Output Set: N:\CRF4\04142006\J522341A.raw

```

106 Ile Ala Ala Gly Asn Ser Ala Asn Leu Ile Ile Leu Pro Ala Glu Asn
107      370      375      380
108 ggg ttt gat gcg ctg cgc cgt cag gtt ccg gta cgt tat tcg gta cgt 1200
109 Gly Phe Asp Ala Leu Arg Arg Gln Val Pro Val Arg Tyr Ser Val Arg
110 385      390      395      400
111 ggc ggc aag gtg att gcc agc aca caa ccg gca caa acc acc gta tat 1248
112 Gly Gly Lys Val Ile Ala Ser Thr Gln Pro Ala Gln Thr Thr Val Tyr
113      405      410      415
114 ctg gag cag cca gaa gcc atc gat tac aaa cgt tga 1284
115 Leu Glu Gln Pro Glu Ala Ile Asp Tyr Lys Arg
116      420      425
119 <210> SEQ ID NO: 2
120 <211> LENGTH: 427
121 <212> TYPE: PRT
122 <213> ORGANISM: Escherichia coli
124 <400> SEQUENCE: 2
125 Val Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly
126 1      5      10      15
127 Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala
128      20      25      30
129 Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp
130      35      40      45
131 Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His
132      50      55      60
133 Leu Asp Thr Thr Gln Thr Ala Gly Gln Pro Asn Trp Asn Gln Ser Gly
134 65      70      75      80
135 Thr Leu Phe Glu Gly Ile Glu Arg Trp Ala Glu Arg Lys Ala Leu Leu
136      85      90      95
137 Thr His Asp Asp Val Lys Gln Arg Ala Trp Gln Thr Leu Lys Trp Gln
138      100      105      110
139 Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp
140      115      120      125
141 Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val
142      130      135      140
143 Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile
144 145      150      155      160
145 Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu
146      165      170      175
147 Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu
148      180      185      190
149 Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr
150      195      200      205
151 Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser
152      210      215      220
153 Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly
154 225      230      235      240
155 Ala Arg Val Thr Ala Ser His Thr Thr Ala Met His Ser Tyr Asn Gly
156      245      250      255
157 Ala Tyr Thr Ser Arg Leu Phe Arg Leu Leu Lys Met Ser Gly Ile Asn

```

RAW SEQUENCE LISTING

DATE: 04/14/2006

PATENT APPLICATION: US/10/522,341A

TIME: 10:12:52

Input Set : F:\Final sequence list-12810-00057-US.txt

Output Set: N:\CRF4\04142006\J522341A.raw

```

158          260          265          270
159 Phe Val Ala Asn Pro Leu Val Asn Ile His Leu Gln Gly Arg Phe Asp
160          275          280          285
161 Thr Tyr Pro Lys Arg Arg Gly Ile Thr Arg Val Lys Glu Met Leu Glu
162          290          295          300
163 Ser Gly Ile Asn Val Cys Phe Gly His Asp Asp Val Phe Asp Pro Trp
164 305          310          315          320
165 Tyr Pro Leu Gly Thr Ala Asn Met Leu Gln Val Leu His Met Gly Leu
166          325          330          335
167 His Val Cys Gln Leu Met Gly Tyr Gly Gln Ile Asn Asp Gly Leu Asn
168          340          345          350
169 Leu Ile Thr His His Ser Ala Arg Thr Leu Asn Leu Gln Asp Tyr Gly
170          355          360          365
171 Ile Ala Ala Gly Asn Ser Ala Asn Leu Ile Ile Leu Pro Ala Glu Asn
172          370          375          380
173 Gly Phe Asp Ala Leu Arg Arg Gln Val Pro Val Arg Tyr Ser Val Arg
174 385          390          395          400
175 Gly Gly Lys Val Ile Ala Ser Thr Gln Pro Ala Gln Thr Thr Val Tyr
176          405          410          415
177 Leu Glu Gln Pro Glu Ala Ile Asp Tyr Lys Arg
178          420          425
181 <210> SEQ ID NO: 3
182 <211> LENGTH: 1284
183 <212> TYPE: DNA
184 <213> ORGANISM: Artificial sequence
186 <220> FEATURE:
187 <223> OTHER INFORMATION: Description of the artificial sequence: coding for
188 cytosine deaminase (codA)
190 <220> FEATURE:
191 <221> NAME/KEY: misc_feature
192 <222> LOCATION: (1)..(3)
193 <223> OTHER INFORMATION: mutation of GTG to ATG start codon for expression
194 in eukaryotic hosts
196 <220> FEATURE:
197 <221> NAME/KEY: CDS
198 <222> LOCATION: (1)..(1281)
199 <223> OTHER INFORMATION: coding for cytosine deaminase (codA)
201 <400> SEQUENCE: 3
202 atg tcg aat aac gct tta caa aca att att aac gcc cgg tta cca ggc 48
203 Met Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly
204 1 5 10 15
205 gaa gag ggg ctg tgg cag att cat ctg cag gac gga aaa atc agc gcc 96
206 Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala
207 20 25 30
208 att gat gcg caa tcc ggc gtg atg ccc ata act gaa aac agc ctg gat 144
209 Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp
210 35 40 45
211 gcc gaa caa ggt tta gtt ata ccg ccg ttt gtg gag cca cat att cac 192
212 Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His

```

RAW SEQUENCE LISTING

DATE: 04/14/2006

PATENT APPLICATION: US/10/522,341A

TIME: 10:12:52

Input Set : F:\Final sequence list-12810-00057-US.txt

Output Set: N:\CRF4\04142006\J522341A.raw

213	50		55		60												
214	ctg	gac	acc	acg	caa	acc	gcc	gga	caa	ccg	aac	tgg	aat	cag	tcc	ggc	240
215	Leu	Asp	Thr	Thr	Gln	Thr	Ala	Gly	Gln	Pro	Asn	Trp	Asn	Gln	Ser	Gly	
216	65					70					75					80	
217	acg	ctg	ttt	gaa	ggc	att	gaa	cgc	tgg	gcc	gag	cgc	aaa	gcg	tta	tta	288
218	Thr	Leu	Phe	Glu	Gly	Ile	Glu	Arg	Trp	Ala	Glu	Arg	Lys	Ala	Leu	Leu	
219					85				90						95		
220	acc	cat	gac	gat	gtg	aaa	caa	cgc	gca	tgg	caa	acg	ctg	aaa	tgg	cag	336
221	Thr	His	Asp	Asp	Val	Lys	Gln	Arg	Ala	Trp	Gln	Thr	Leu	Lys	Trp	Gln	
222					100				105						110		
223	att	gcc	aac	ggc	att	cag	cat	gtg	cgt	acc	cat	gtc	gat	gtt	tcg	gat	384
224	Ile	Ala	Asn	Gly	Ile	Gln	His	Val	Arg	Thr	His	Val	Asp	Val	Ser	Asp	
225			115					120					125				
226	gca	acg	cta	act	gcg	ctg	aaa	gca	atg	ctg	gaa	gtg	aag	cag	gaa	gtc	432
227	Ala	Thr	Leu	Thr	Ala	Leu	Lys	Ala	Met	Leu	Glu	Val	Lys	Gln	Glu	Val	
228		130					135					140					
229	gcg	ccg	tgg	att	gat	ctg	caa	atc	gtc	gcc	ttc	cct	cag	gaa	ggg	att	480
230	Ala	Pro	Trp	Ile	Asp	Leu	Gln	Ile	Val	Ala	Phe	Pro	Gln	Glu	Gly	Ile	
231	145					150				155					160		
232	ttg	tcg	tat	ccc	aac	ggg	gaa	gcg	ttg	ctg	gaa	gag	gcg	tta	cgc	tta	528
233	Leu	Ser	Tyr	Pro	Asn	Gly	Glu	Ala	Leu	Leu	Glu	Glu	Ala	Leu	Arg	Leu	
234				165					170						175		
235	ggg	gca	gat	gta	gtg	ggg	gcg	att	ccg	cat	ttt	gaa	ttt	acc	cgt	gaa	576
236	Gly	Ala	Asp	Val	Val	Gly	Ala	Ile	Pro	His	Phe	Glu	Phe	Thr	Arg	Glu	
237				180					185					190			
238	tac	ggc	gtg	gag	tcg	ctg	cat	aaa	acc	ttc	gcc	ctg	gcg	caa	aaa	tac	624
239	Tyr	Gly	Val	Glu	Ser	Leu	His	Lys	Thr	Phe	Ala	Leu	Ala	Gln	Lys	Tyr	
240			195					200					205				
241	gac	cgt	ctc	atc	gac	gtt	cac	tgt	gat	gag	atc	gat	gac	gag	cag	tcg	672
242	Asp	Arg	Leu	Ile	Asp	Val	His	Cys	Asp	Glu	Ile	Asp	Asp	Glu	Gln	Ser	
243		210					215					220					
244	cgc	ttt	gtc	gaa	acc	gtt	gct	gcc	ctg	gcg	cac	cat	gaa	ggc	atg	ggc	720
245	Arg	Phe	Val	Glu	Thr	Val	Ala	Ala	Leu	Ala	His	His	Glu	Gly	Met	Gly	
246	225					230				235					240		
247	gcg	cga	gtc	acc	gcc	agc	cac	acc	acg	gca	atg	cac	tcc	tat	aac	ggg	768
248	Ala	Arg	Val	Thr	Ala	Ser	His	Thr	Thr	Ala	Met	His	Ser	Tyr	Asn	Gly	
249				245					250					255			
250	gcg	tat	acc	tca	cgc	ctg	ttc	cgc	ttg	ctg	aaa	atg	tcc	ggg	att	aac	816
251	Ala	Tyr	Thr	Ser	Arg	Leu	Phe	Arg	Leu	Leu	Lys	Met	Ser	Gly	Ile	Asn	
252				260					265				270				
253	ttt	gtc	gcc	aac	ccg	ctg	gtc	aat	att	cat	ctg	caa	gga	cgt	ttc	gat	864
254	Phe	Val	Ala	Asn	Pro	Leu	Val	Asn	Ile	His	Leu	Gln	Gly	Arg	Phe	Asp	
255			275					280				285					
256	acg	tat	cca	aaa	cgt	cgc	ggc	atc	acg	cgc	gtt	aaa	gag	atg	ctg	gag	912
257	Thr	Tyr	Pro	Lys	Arg	Arg	Gly	Ile	Thr	Arg	Val	Lys	Glu	Met	Leu	Glu	
258		290					295					300					
259	tcc	ggc	att	aac	gtc	tgc	ttt	ggg	cac	gat	gat	gtc	ttc	gat	ccg	tgg	960
260	Ser	Gly	Ile	Asn	Val	Cys	Phe	Gly	His	Asp	Asp	Val	Phe	Asp	Pro	Trp	
261	305					310				315					320		

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 04/14/2006
PATENT APPLICATION: US/10/522,341A TIME: 10:12:53

Input Set : F:\Final sequence list-12810-00057-US.txt
Output Set: N:\CRF4\04142006\J522341A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:55; N Pos. 5014
Seq#:57; N Pos. 6697
Seq#:72; Xaa Pos. 2,7,9,11
Seq#:73; Xaa Pos. 9,11
Seq#:74; Xaa Pos. 7
Seq#:75; Xaa Pos. 6
Seq#:77; Xaa Pos. 5
Seq#:79; Xaa Pos. 6
Seq#:86; N Pos. 25,26,27,28,29,30,31,32,33,34,35,36,37,38,39
Seq#:87; N Pos. 25,26,27,28,29,30,31,32,33,34,35,36,37,38,39
Seq#:130; N Pos. 25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43
Seq#:130; N Pos. 44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62
Seq#:130; N Pos. 63
Seq#:163; N Pos. 6,12
Seq#:164; N Pos. 6,12
Seq#:165; N Pos. 6,12
Seq#:166; N Pos. 6,12
Seq#:167; N Pos. 6,12
Seq#:168; N Pos. 6,12
Seq#:174; N Pos. 25,26,27,28,29,30,31,32,33,34,35,36,37,38
Seq#:175; N Pos. 25,26,27,28,29,30,31,32,33,34,35,36,37,38

VERIFICATION SUMMARY

DATE: 04/14/2006

PATENT APPLICATION: US/10/522,341A

TIME: 10:12:53

Input Set : F:\Final sequence list-12810-00057-US.txt

Output Set: N:\CRF4\04142006\J522341A.raw

L:3928 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:55 after pos.:4980
L:4305 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:57 after pos.:6660
L:5128 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:72 after pos.:0
L:5153 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73 after pos.:0
L:5173 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:74 after pos.:0
L:5193 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:75 after pos.:0
L:5227 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:77 after pos.:0
L:5261 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:79 after pos.:0
L:5358 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:86 after pos.:0
L:5376 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:87 after pos.:0
L:5898 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:130 after pos.:0
L:5900 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:130 after pos.:60
L:6307 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:163 after pos.:0
L:6330 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:164 after pos.:0
L:6353 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:165 after pos.:0
L:6376 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:166 after pos.:0
L:6399 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:167 after pos.:0
L:6422 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:168 after pos.:0
L:6500 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:174 after pos.:0
L:6518 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:175 after pos.:0